ADVANCES IN (STANDARDIZATION OF) ORGANOMETALLIC AND INORGANIC STRUCTURE REPRESENTATIONS

Roger Sayle and John Mayfield
NextMove Software, Cambridge, UK
Greg Landrum
ETH Zurich, Zurich, Switzerland

ACS Spring 2022 Meeting, CINF Division, San Diego/Zoom, Tuesday 22nd March 2022

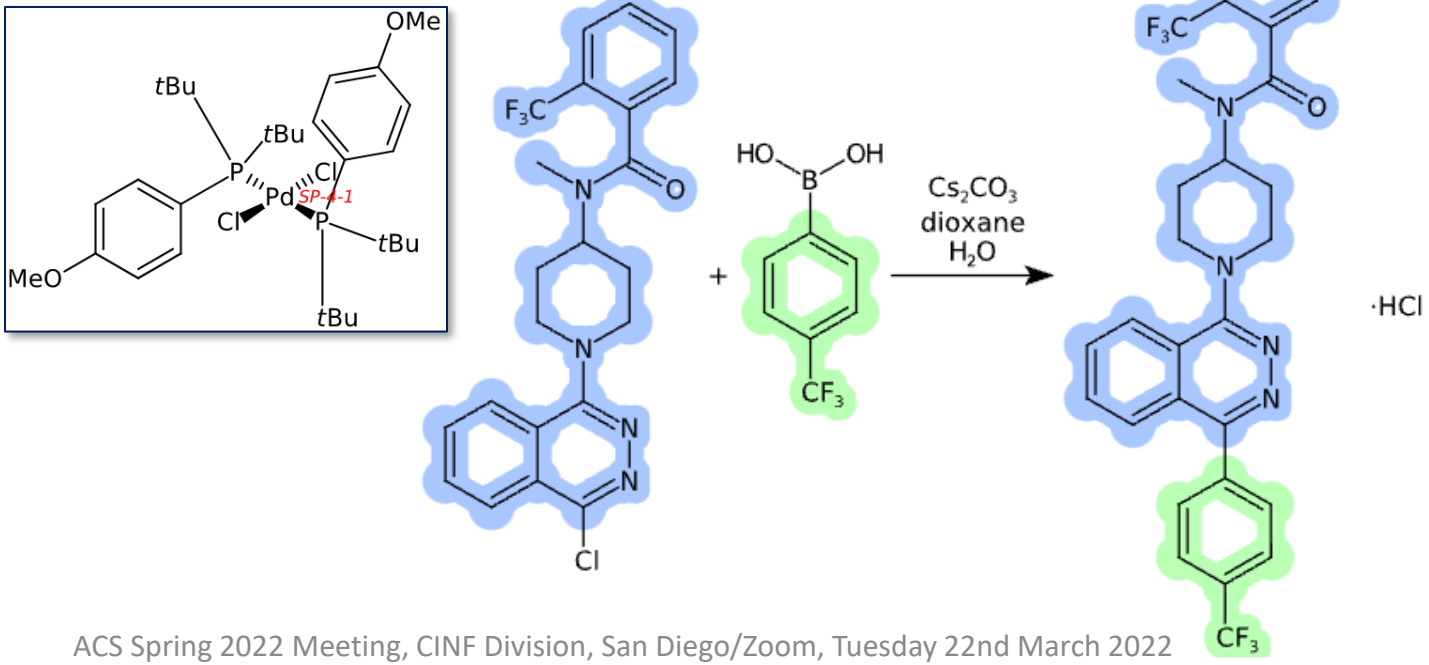
MOTIVATIONS

- Representing reagents and catalysts in reactions.
- Lack of support/normalization in ChEMBL/PubChem.
- Slow progress from InChI “inorganic” working group.
- Contribute/influence proof-of-concept in RDKit.

US20110046143A1 [0102]

N-Methyl-2-(trifluoromethyl)-N-(1-(4-(4-(trifluoromethyl)phenyl)phthalazin-1-yl) piperidin-4-yl)benzamide hydrochloride

Charge a microwave vessel with N-(1-(4-chlorophthalazin-1-yl)piperidin-4-yl)-N-methyl-2-(trifluoromethyl) benzamide (0.101 g, 0.23 mmol), 4-(trifluoromethyl)phenylboronic acid (0.171 g, 0,9 mmol), cesium carbonate (0.295 g, 0.91 mmol), 1,4-dioxane (3 mL), and water (1 mL). Purge the reaction vial two times with nitrogen. Add **(SP-4-1)-bis[bis(1,1-dimethylethyl)(4-methoxyphenyl)phosphine-κP]dichloro-palladium** (J. Org. Chem. 2007, 72, 5104-5112) (0.002 g; 0.003 mmol) and heat the reaction at 90° C. for 16 h. After cooling, separate the two layers and remove the water. Evaporate the organic solvent with a stream of nitrogen. Purify the residue from the organic layer using silica gel chromatography (0-10% methanol in dichloromethane). Add 4 N HCl in dioxane to a solution of the isolated product in methanol and remove the solvents in vacuo to obtain the title compound (0.100 g, 75%). ES/MS m/z 559.2 (M+1).



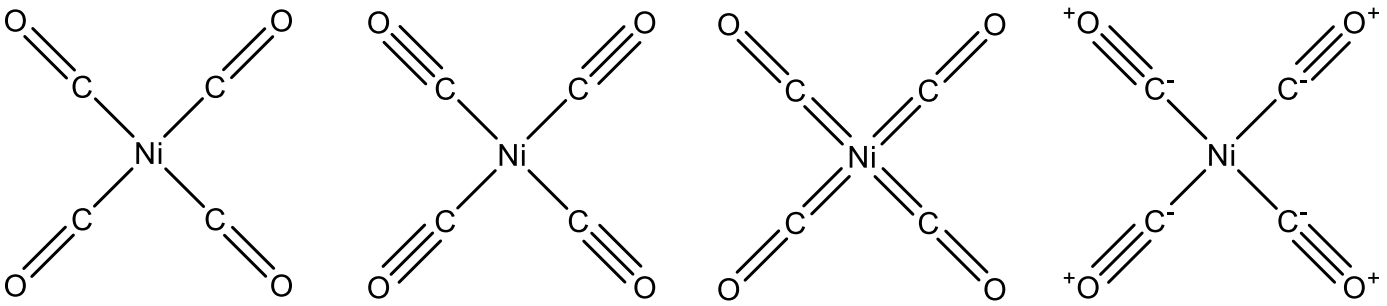
ALEX CLARK'S 4 RULES OF INORGANICS

- https://github.com/aclarkxyz/data_coordinchi

1. The heavy atom graph must be "complete".
2. Hydrogen counts must be known unambiguously.
3. Bond orders descriptive enough to capture delocalization.
4. Formal charges must be placed where the can be balanced.

EXAMPLE: CARBONYL NORMALIZATION

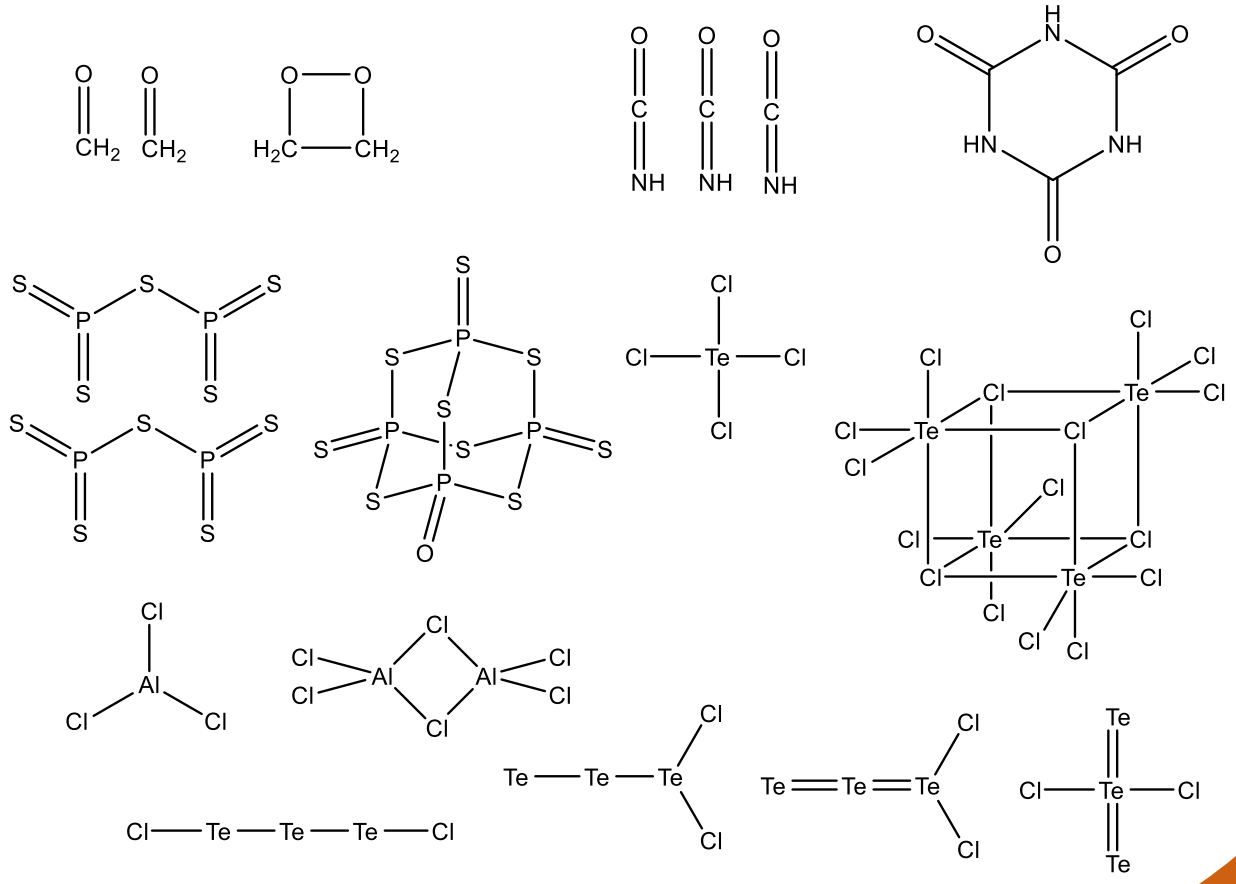
- Chemists use a variety of representations of metal carbonyls.



- All of these representations can/should be considered equivalent. *[C]=O, *C#O, *=C=O, *[C-]#[O+], etc.
- Additional [M-]C#[O+], [M+][C-]=O, [M-][C+]=O and [M+]#C[O-] are equivalent under electron migration.
- Unfortunately, a form commonly deposited in PubChem is *[CH]=O, which has an additional hydrogen.

THE CHALLENGE OF COMPLETENESS

- Dimerization, Trimerization and polymerization.

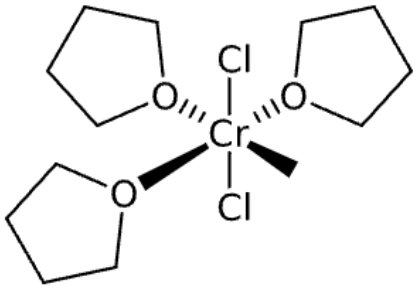


SOLUTION: EMPIRICAL FORMULA

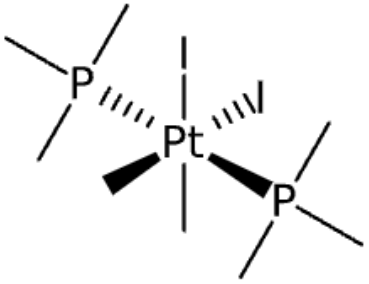
- The solution to “completeness” is the use of empirical molecular formula, where the count of each element is divided by their greatest common divisor (GCD): Hence Te_4Cl_{16} becomes TCl_4.
- MolHash technology was described at the last San Diego ACS meeting in 2019, contributed to RDKit.
  – https://www.nextmovesoftware.com/talks/OBoyle_MolHash_ACS_201908.pdf
- Efficient GCD algorithms can use bit sets of unique prime factors, e.g. for CHEMBL555389:
  – $C_{156}H_{192}N_{27}O_{90}P_{33}S_6^{-24}$ -> $C_{52}H_{64}N_9O_{30}P_{11}S_2^{-8}$

NON-TETRAHEDRAL STEREOCHEMISTRY

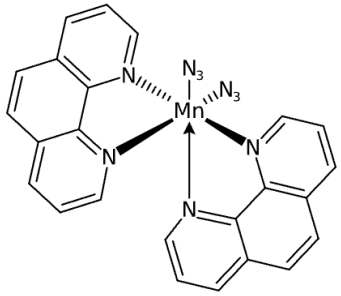
- The significant representational challenge with organometallic and inorganic structures is non-tetrahedral stereochemistry/chirality.
- Octahedral examples:



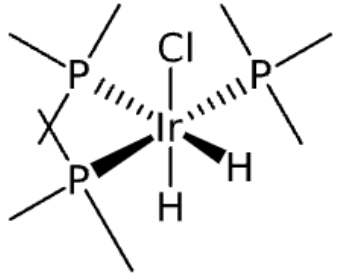
COD4073210



COD4130638



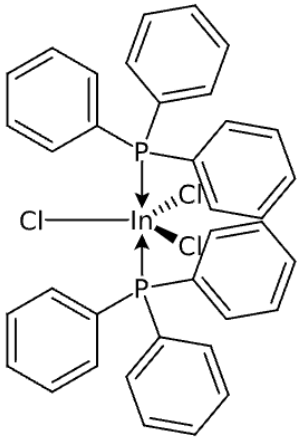
COD2008644



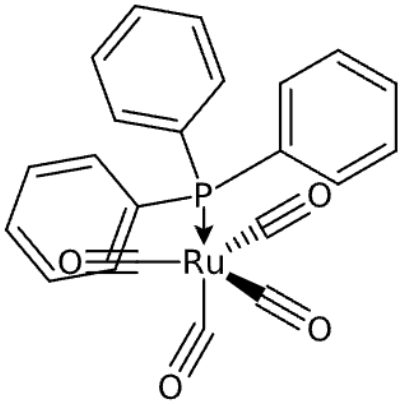
COD4078254

- Structures from 3D co-ordinates in the COD database, depictions by CDK/Beam.

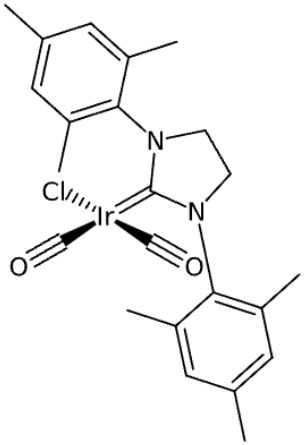
NON-OCTAHEDRAL EXAMPLES



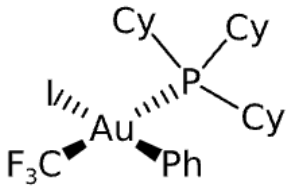
COD4102282



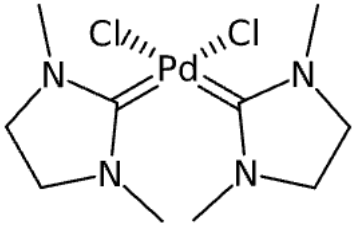
COD4334746



COD4070781



COD4121836



COD1514811

RECENT PROGRESS

- A set of proposals was presented at the 2021 RDKit UGM in October 2021.
- Most of the core functionality is now in RDKit (on a branch) and should be generally available in the next major release 2022.09.
  - https://github.com/rdkit/rdkit/issues/4851
  - https://github.com/rdkit/rdkit/pull/5084
- The rest of this talk summarizes what's in place, remaining work in progress and some additional suggestions.

A STEREOCHEMISTRY ONTOLOGY

- Beyond tetrahedral stereochemistry and its two permutations (parity), are a variety of geometries with different (larger) numbers of permutations.

Class	Degree	Pairs	IUPAC	SMILES	Perm	Parent
Octahedral	6	3	(OC-6)	@OH	30	
Trigonal bipyramidal	5	1	(TBPY-5)	@TB	20	
Square planar	4	2	(SP-4)	@SP	3	@OH
T-shape	3	1	(TS-3)		3	@SP/@OH
Square pyramid	5	2	(SPY-5)		30	@OH
See-saw (major)	4	1	(SS-4)		20	@TB
See-saw (minor)	4	1	(SS-4)			@OH

THE IMPORTANCE OF LINEAR PAIRS

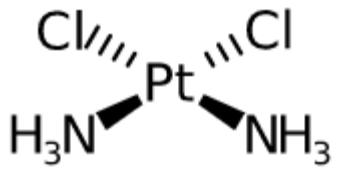
- A useful conceptual abstraction are pairs of neighbours that are linearly across from one another.
- The number of pairs present is a defining characteristic of a non-tetrahedral chirality class.
- The permutation index encodes the identities of these pairs and (typically) the parity/handedness.
  - Permutation = Pairs + Parity.
- These pairs (and parity) need to be preserved when interconverting between 3D, 2D and 1D [i.e. SMILES].

INTERNAL REPRESENTATION

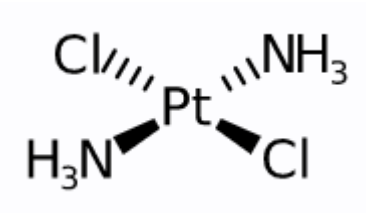
- A cheminformatics toolkit need only keep track of two extra pieces of information to track advanced stereochemistry: the chirality class (e.g. tetrahedral, square planar, trigonal bipyramidal, octahedral), and the permutation number, 1..N.
- Technically, both can/could be encoded in a single integer value.
- One proposal detailed in a few slides, proposes making the permutation optional (encoded as zero).
- SMILES permutations are conveniently documented.

SP-4 EXAMPLES

- Let's start with a classic example:



CHEBI27899
cisplatin
(SP-4-2)-diamminedichloroplatinum
Cl[Pt@SP1](Cl)([NH3])[NH3]



CHEBI35852
transplatin
(SP-4-1)-diamminedichloroplatinum
Cl[Pt@SP2](Cl)([NH3])[NH3]

WORKED EXAMPLE(S)

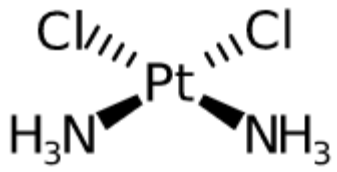
P[M@SP1](Q)(R)S means P is across from R, Q across from S. “U”

P[M@SP2](Q)(R)S means P is across from Q, R across from S. “4”

P[M@SP3](Q)(R)S means P is across from S, Q across from R. “Z”

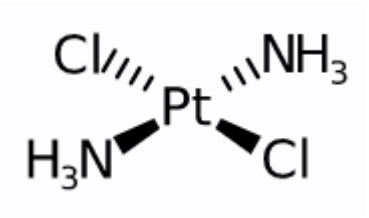
Cl[Pt@SP1](Cl)([NH3])[NH3]

Cl[Pt@SP3](Cl)([NH3])[NH3]



cisplatin

Cl[Pt@SP2](Cl)([NH3])[NH3]



transplatin

See the excellent documentation at http://opensmiles.org

API FUNCTIONS FOR STEREOCHEMISTRY

- To simplify working with permutation numbers, RDKit now supports a set of “helper” functions:

```
bool Atom::invertChirality();
Bond *getChiralAcrossBond(Atom *cen, Bond *qry);
Bond *getChiralAcrossBond(Atom *cen, Atom *qry);
Atom *getChiralAcrossAtom(Atom *cen, Bond *qry);
Atom *getChiralAcrossAtom(Atom*cen, Atom *qry);
Bond *getTrigonalBipyramidalAxialBond(Atom *cen, int which=0);
Atom *getTrigonalBipyramidalAxialBond(Atom *cen, int which=0);
int isTrigonalBipyramidalAxialBond(Atom *cen, Bond *qry);
int isTrigonalBipyramidalAxialAtom(Atom *cen, Atom *qry);
```

RDKIT STATE-OF-THE-UNION

- Internal Representation
- From 3D co-ordinates
- From SMILES
- From Mol file/sketch
- From IUPAC/CIP name

- To 3D co-ordinates
- To SMILES
- To Mol file/sketch
- To IUPAC/CIP name

RDKIT STATE-OF-THE-UNION

- Internal Representation
- From 3D co-ordinates
- From SMILES
- From Mol file/sketch
- From IUPAC/CIP name

- To 3D co-ordinates
- To SMILES
- To Mol file/sketch
- To IUPAC/CIP name

Fully implemented

Mostly implemented/More details here

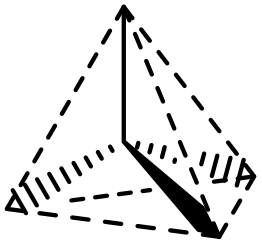
Not yet/More details here

SCIENCE STATE-OF-THE-ART

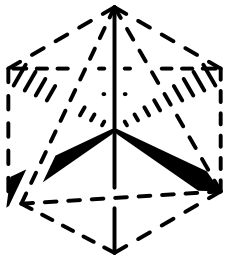
- Level 1. Distinct Stereochemistry
  - All neighbors (constitutionally) distinct.
- Level 2. Symmetry Degeneracy
  - Some neighbors (constitutionally) equivalent.
- Level 3. Geometry-only specification
  - Geometry known, but permutation unspecified.
- Level 4. Partial co-ordination (SQPY, TPY, SS, T).
  - Neighbours include lone pairs and/or implicit hydrogens.

- Daylight 1..2, IUPAC 1..4, CDK 1..2, Current RDKit 1→2/3/4.
- John Mayfield quote: “It’s useful, but it’s not SMILES”.

TETRAHEDRAL VOLUME CONSTRAINTS



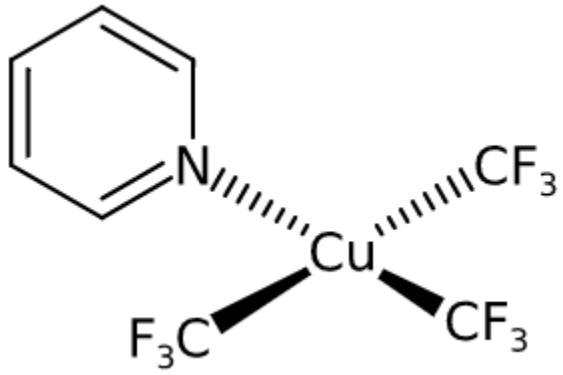
- When generating 3D co-ordinates with distance geometry, one, four or five tetrahedral volume constraints are added for each chiral tetrahedral centre.



- For octahedral centers, eight tetrahedra are required, for trigonal bipyramidal six are required, and for see-saw two are required.
- $V = \frac{|(a-d)\cdot((b-d)\times(c-d))|}{6}$, ignoring modulus for signed volume.

LEVEL 3: EXTENSION TO GEOMETRY

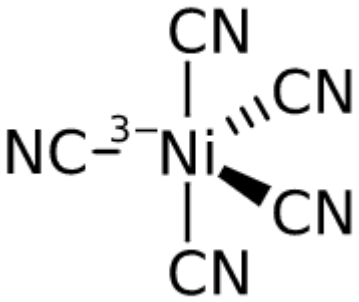
- It's useful to annotate symmetric centres as @SP.



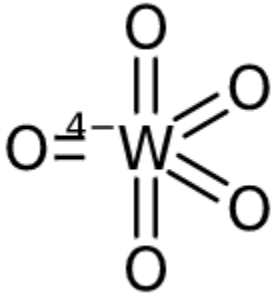
COD4128444
(SP-4)-(pyridine)tris(trifluoromethyl)copper
[Cu@SP]([N]1=CC=CC=C1)(C(F)(F)F)(C(F)(F)F)C(F)(F)F

LEVEL 3: EXTENSION TO GEOMETRY

- It's useful to annotate symmetric centres as @TB.



CHEBI30370
(TBPY-5)-pentacyanonickelate(3-)
N#C[Ni@TB-3](C#N)(C#N)(C#N)C#N



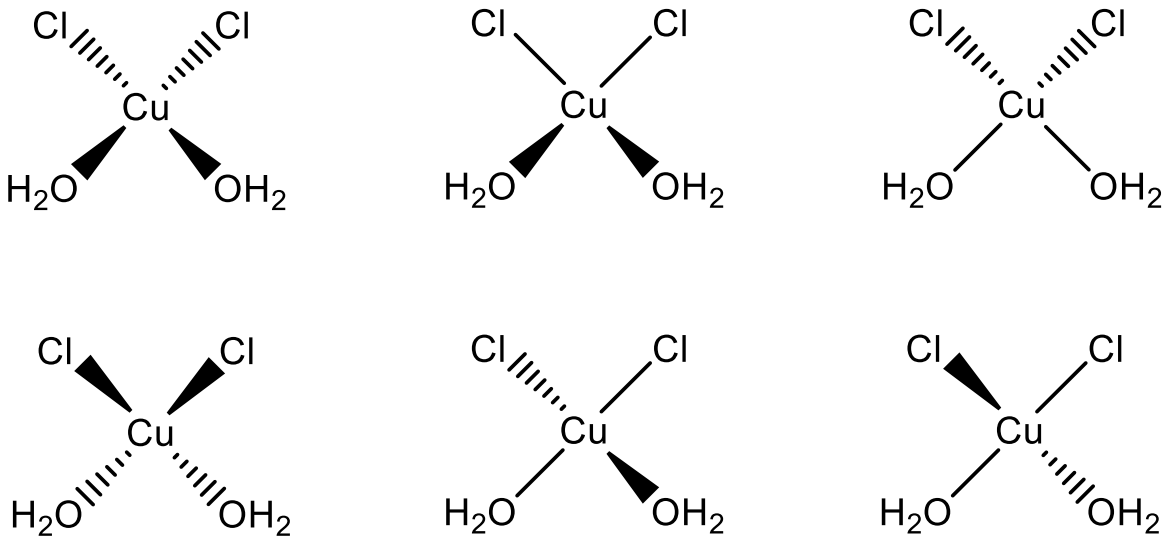
CHEBI30370
(TBPY-5)-pentaoxotungstate(4-)
O=[W@TB-4](=O)(=O)(=O)=O

LEVEL 3: EXTENSION TO GEOMETRY

- Sometimes 4-valent carbon isn't tetrahedral!
- [Li][C@SP]([H])([H])[Li]
- [Li][C@SPH2][Li]

SQUARE PLANAR IN SKETCHES

- Square planar centres require two pairs of linear bonds, in any of the following configurations.



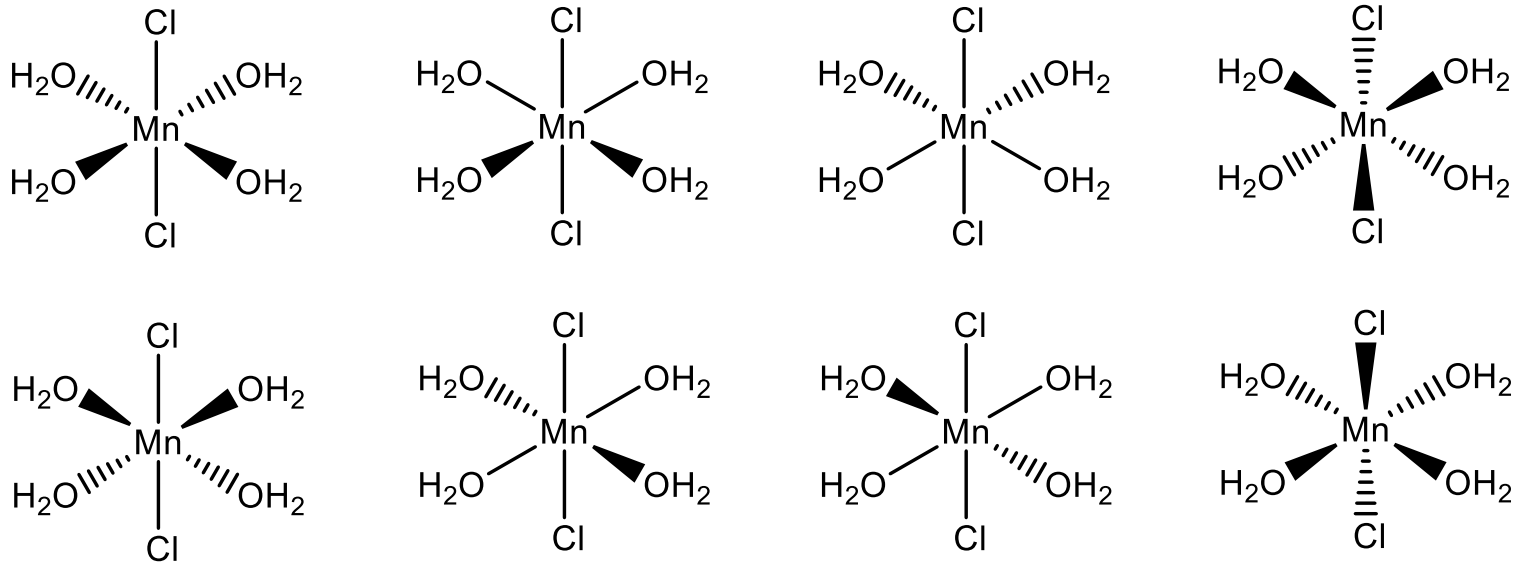
* This conflicts with ST-2.6 of IUPAC's Graphical Representation of Stereochemical Configuration 2006

DEPICTION/PERCEPTION ALGORITHMS

- When neighbors are ordered clockwise, the permutation is always SP3 (i.e. “U” form).
- Valid input wedge (W)/hash (H) assignments are:
  - WHHW, HHWW, HWWH, WWHH, WNHN, NHNW, HNWN, NWNH,NWWN, WWNN, WNNW, NNWW, NHHN, HHNN, HNNH, NNHH (i.e. rotationally invariant).
- Output assignment can be arbitrary, i.e. WHHW.

OCTAHEDRAL STEREO IN SKETCHES

- Octahedral centres require three pairs of linear bonds, in any of the following configurations.

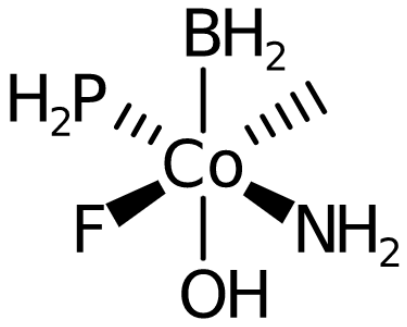


* This conflicts with ST-2.11 of IUPAC's Graphical Representation of Stereochemical Configuration 2006

DEPICTION/PERCEPTION ALGORITHMS

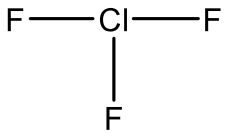
- When the six neighbors are ordered clockwise, the permutation is always either OH6 or OH18.
- NHWNWH is OH18.
- Swapping any axial pair (a,d), (b,e) or (c,f) inverts chirality.

- [Co@OH18](B)(C)(N)(O)(F)P

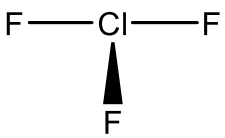


T-SHAPE STEREO IN SKETCHES

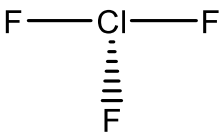
- Limit non-tetrahedral stereochemistry to elements with atomic numbers 15 or above.
- Non-tetrahedral stereochemistry requires at least one pair of bonds to be (close to) linear.
- From 2D, stereochemistry is only perceived on atoms with explicit wedge/hash bonds*.



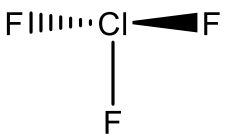
Unspecified



T-Shape



T-Shape



T-Shape

* This conflicts with ST-2.3 of IUPAC's Graphical Representation of Stereochemical Configuration 2006

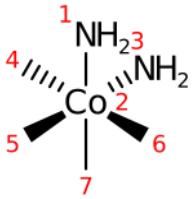
WHERE AND WHY DO WE DIFFER?

- All “preferred” IUPAC configurations are understood.
- These rules are affine (i.e. rotationally) invariant.
- The mirror image of the 2D depiction represents the mirror image of the 3D structure.
- This typically provides interpretation for IUPAC’s “unacceptable” sketch representations, for which no semantic interpretation is given.
- Existing 2D depictions retain their meaning, and non-tetrahedral geometry is only interpreted in cases where tetrahedral chemistry would be invalid.

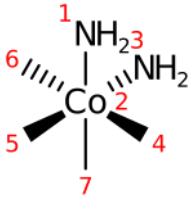
CANONICAL CONFIGURATION

Canonical configurations are more tricky than **tetrahedral** due to symmetry. These all mean the same**:**

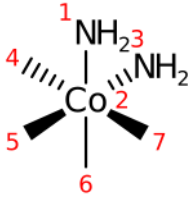
Notice symmetric atoms are moving around in layout for different config



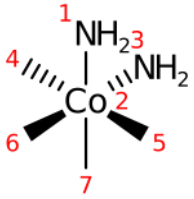
`N[Co@OH1](N)(C)(C)(C)C`



`N[Co@OH2](N)(C)(C)(C)C`



`N[Co@OH3](N)(C)(C)(C)C`



`N[Co@OH4](N)(C)(C)(C)C`

*Select the **lowest number** one as canonical, how?*

WRITING CANONICAL SMILES

- Current RDKit makes use of clever tables to capture how permutations are updated when swapping neighbors X and Y.
- This allows non-canonical SMILES to be written as the reordering of atoms is tracked during SMILES' depth-first traversal.
- The missing insight is that these same swap/exchange tables can be used in combination with symmetry data. If neighbors X and Y share the same symmetry class, if the "swapped" permutation is lower use that.

PSEUDO-CODE

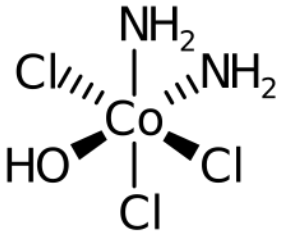
```
perm_set = [ orig_perm ]
for x in nbors:
  for y in nbors where sym(x) == sym(y) and x != y:
    new_perms = [swap(p,x,y) for p in perm_set]
    perm_set += new_perms
canon_perm = min(perm_set)
```

- For example, if all nbors are in the same symclass, canon_perm should be 1, for any value of orig_perm.
- A uint32 (bitmap) can be used to hold the perm_set.

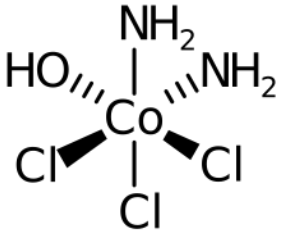
CANONICAL CONFIGURATION

```
Cl[Co@OH1](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH2](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH3](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH4](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH5](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH6](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH7](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH8](N)(N)(O)(Cl)Cl => N[Co@OH25](N)(O)(Cl)(Cl)Cl
Cl[Co@OH9](N)(N)(O)(Cl)Cl => N[Co@OH25](N)(O)(Cl)(Cl)Cl
Cl[Co@OH10](N)(N)(O)(Cl)Cl => N[Co@OH25](N)(O)(Cl)(Cl)Cl
Cl[Co@OH11](N)(N)(O)(Cl)Cl => N[Co@OH25](N)(O)(Cl)(Cl)Cl
Cl[Co@OH12](N)(N)(O)(Cl)Cl => N[Co@OH25](N)(O)(Cl)(Cl)Cl
Cl[Co@OH13](N)(N)(O)(Cl)Cl => N[Co@OH25](N)(O)(Cl)(Cl)Cl
Cl[Co@OH14](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH15](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH16](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH17](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH18](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH19](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH20](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH21](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH22](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH23](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH24](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH25](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH26](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH27](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH28](N)(N)(O)(Cl)Cl => N[Co@OH8](N)(O)(Cl)(Cl)Cl
Cl[Co@OH29](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
Cl[Co@OH30](N)(N)(O)(Cl)Cl => N[Co@OH1](N)(O)(Cl)(Cl)Cl
```

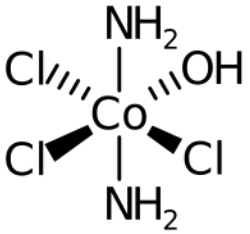
3 configurations



@OH8



@OH1

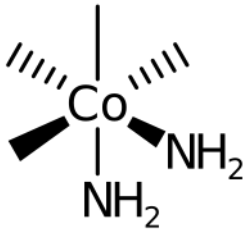


@OH25

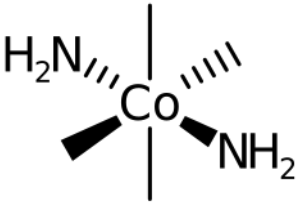
CANONICAL CONFIGURATION

```
N[Co@OH1](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH2](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH3](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH4](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH5](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH6](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH7](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH8](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH9](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH10](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH11](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH12](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH13](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH14](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH15](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH16](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH17](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH18](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH19](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH20](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH21](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH22](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH23](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH24](N)(C)(C)(C)C => C[Co@OH1](C)(C)(C)(N)N
N[Co@OH25](N)(C)(C)(C)C => C[Co@OH12](C)(C)(C)(N)N
N[Co@OH26](N)(C)(C)(C)C => C[Co@OH12](C)(C)(C)(N)N
N[Co@OH27](N)(C)(C)(C)C => C[Co@OH12](C)(C)(C)(N)N
N[Co@OH28](N)(C)(C)(C)C => C[Co@OH12](C)(C)(C)(N)N
N[Co@OH29](N)(C)(C)(C)C => C[Co@OH12](C)(C)(C)(N)N
N[Co@OH30](N)(C)(C)(C)C => C[Co@OH12](C)(C)(C)(N)N
```

2 configurations, cis/trans-



@OH1

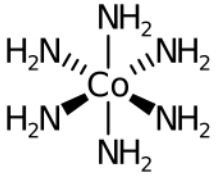


@OH12

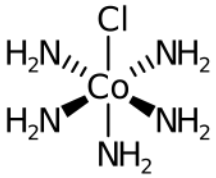
REDUNDANT CONFIGURATION?

Octahedral specification is **redundant** if all possible configurations @OH1…@OH30 are automorphic

`N[Co@OH1](N)(N)(N)(N)N`



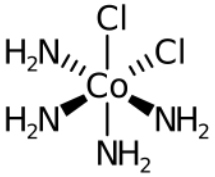
✖



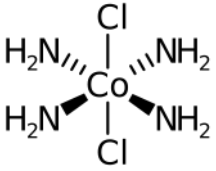
`Cl[Co@OH1](N)(N)(N)(N)N`

✖

`Cl[Co@OH1](Cl)(N)(N)(N)N`



✔ *cis-*



`Cl[Co@OH25](Cl)(N)(N)(N)N`

✔ *trans-*

LEVEL 4: INTO THE GREAT UNKNOWN

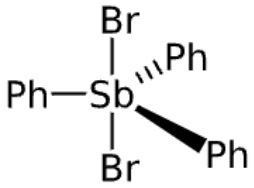
- It would be nice to handle partial co-ordination geometries, square pyramidal is octahedral with a lone pair, trigonal pyramidal is trigonal bipyramidal with an axial lone pair, T-shape is square planar with a lone pair, etc.
- Unfortunately, Daylight never implemented these cases and offered no (little) guidance on how to adapt permutation numbers/indices.
- Consensus seems to be that unfilled valences appear use the parent idx/position, Hs before LPs.

SMARTS MATCHING

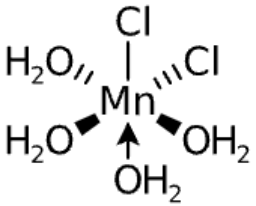
- [*@OH] can be used to identify octahedral centers.
- *[@OH1]* can be used to identify “across” atoms, and *[@OH3]* can be used to identify cis atoms.
- SMARTS matchers traditionally loop over all unvisited neighbors to extend the match by the next atom.
- At non-tetrahedral stereo centers, the matcher can efficiently use the proposed getChiralAcrossAtom.
- Like regular tetrahedral matching, once three axes (neighbours) have be identified, the parity can be checked.

TRADITIONAL NAMING

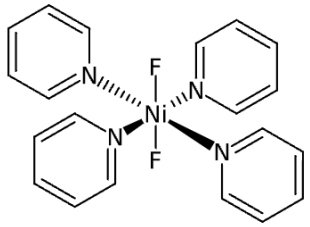
- Many non-tetrahedral centres may be named using cis/trans.



COD2007221
trans-dibromidotriphenyl antimony

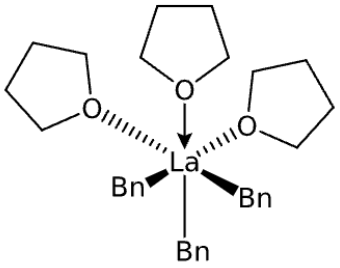


COD8103034
cis-tetraaquadichlorido manganese

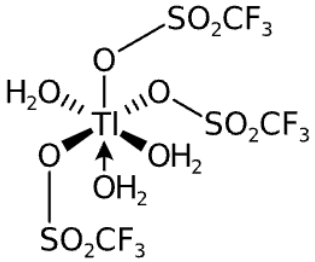


COD7035539
trans-diflurotetrakis (pyridine)nickel

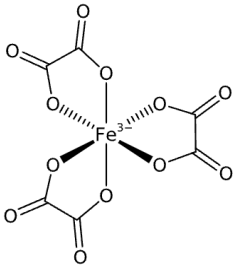
- Others using fac-, mer-, Delta- (Δ-) and Lambda (Λ-).



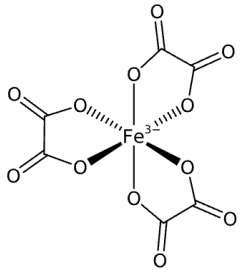
COD4075390
fac-tribenzyltris(tetrahydrofuran) lanthanum



COD7011937
mer-triaquatris(triflato-κO) thallium



Δ-tris(oxalate) ferrate(3-)



Λ-tris(oxalate) ferrate(3-)

IUPAC NAMING

- IUPAC's blue book contains recommendations for stereochemical prefixes based on CIP priority rules.
- See "Algorithmic Analysis of Cahn-Ingold-Prelog Rules of Stereochemistry: Proposals for Revised Rules and a Guide for Machine Implementation", JCIM 2018.

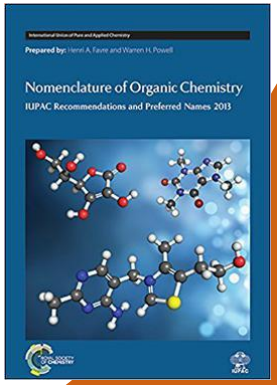
(OC-6-12)-dichlorobis(4-amino-1,2,4-triazole-κN1)dimethyltin

(TBPY-5-12)-(tricyclohexylphosphine)tetracarbonyliron

(OC-6-43)-bis(triphenylphosphine)dichloridohydridooxidorhenium

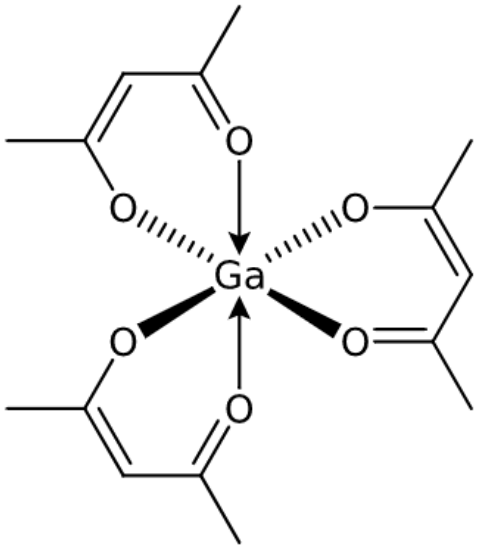
(SP-4-3)-(tricyclohexylphosphine)carbonylchloridoiodidopalladium

(SPY-5-35-C)-5-phenyl-1-oxa-thia-5λ^5-phosphaspiro[4.4]nonane



PUSHING THE LIMITS...

- Unfortunately, CIP's priority rules can't (yet) handle resonant ligands in difficult cases.



- COD1548616 Δ-tris(acetoacetato)gallium

THE PROOF OF THE PUDDING

- One motivation for this work was the lack of a inorganic stereochemistry data set at InChI workshop.
- Nearly 100 examples (from COD) were contributed to RDKit's testsuite.
- Using RDKit directly, a database of around 65231 SMILES with non-tetrahedral stereochemistry can be perceived from COD (via Data Warrior).
  - 38434 octahedral examples
  - 21540 square planar examples
  - 8494 trigonal bipyramidal examples

THE FLY IN THE OINTMENT

- Frustratingly an initial analysis of this data set reveals the inorganic standardization work still left to be done.
- Example: COD1000020
- RDKit SMILES from Data Warrior V2000 export:
  - CC[As](CC)(CC)~[Pt@SP2+2](~[Cl-])(~[Cl-])~[As](CC)(CC)CC
- RDKit SMILES from Data Warrior V3000 export:
  - CC[AsH](CC)(CC)<-[Pt@SP2+2](->[Cl-])(->[Cl-])->[AsH](CC)(CC)CC
- Unfortunately the non-standard bond orders (and unconstrained valences) used by Data Warrior are misunderstood by RDKit [which itself differs from Biovia's interpretation of the same files].

SUMMARY & CONCLUSIONS

- More progress is being made in representation of organometallic and inorganic compounds in cheminformatics.

ACKNOWLEDGEMENTS

- John Mayfield, NextMove Software
- Greg Landrum, ETH Zurich/T5 informatics

- David Weininger
- Daniel Lowe
- Ian Bruno
- Alex Clark
- Colin Batchelor
- Hinnerk Rey
- Jonathan Brecher
- Jonathan Goodman
- Evan Bolton
- Matt Swain